Regulating gene expression

Jansen, Ritsert C.; Nap, Jan-Peter

Published in:
Trends in Genetics

DOI:
10.1016/j.tig.2004.03.002

Publication date:
2004

Update TRENDS in Genetics Vol.20 No.5 May 2004

Research Focus

Regulating gene expression: surprises still in store

Ritsert C. Jansen[1] and Jan-Peter Nap[1,2,3]

[1]Groningen Bioinformatics Centre (GBIC), University of Groningen (RUG), PO Box 800, NL-9700AV Groningen, The Netherlands
[2]Bioinformatics, Hanzehogeschool Groningen, Zernikeplein 11, NL-9747AS Groningen, The Netherlands
[3]Applied Bioinformatics, Plant Research International, PO Box 16, NL-6700AA Wageningen, The Netherlands

Understanding how genes constitute and contribute to the regulatory networks that result in phenotypic diversity is the major challenge of the post-genome era. Recently, it has been shown that major players in gene regulation can be identified by genome-wide linkage analysis of whole-genome gene expression profiles that were obtained in segregants from yeast strains. Surprisingly, these important elements in regulatory variation are dispersed across several gene ontology classes. Therefore, the regulatory role of transcription factors seems less than was previously anticipated.

Significant differences in gene expression between members of a population can be caused by genetic variation between individuals within that population. The variants of a particular gene or its upstream region can introduce quantitative variation in gene expression '*in cis*' in several ways: (i) altering functional motifs in the promoter region; (ii) changing the stability of the mRNA; or (iii) modifying the gene product in such a way that the feedback control of transcription is shifted. Gene variants can also introduce quantitative variation in the expression of other genes '*in trans*'. The mechanisms of such *trans*-acting regulatory variation are largely unknown. Unanswered questions include: what types of genes are involved in *trans*-linkage? Are polymorphisms of transcription factors an important factor in regulatory variation or are other classes of genes equally or more important?

A recent study by Yvert and coworkers [1] provides new insights into these issues, extending and elaborating on their earlier analysis of transcriptional regulation in yeast [2]. At the centre of this approach was the analysis of genetic variation that occurred when a yeast laboratory strain (BY4716) was crossed with a yeast wild strain (RM11-1a). This gave rise to a segregating population of 86 genetically different progeny, traditionally seen as a 'mapping population'. Genotyping such a population, using molecular markers as flags of DNA sequence differences, generates a compositional linkage map of the essentially Mendelian assortment of genes that occurs after crossing. Genome-wide expression profiling using microarrays provides a picture of variations in gene expression in the population. Yvert and coworkers combined genotyping and expression profiling to study gene regulation [1]. When the expression of a particular gene in the population differs between the individuals, it can be considered an 'expression phenotype'. For such a phenotype, linkage analysis is as appropriate as it is for 'classical' phenotypes or traits, such as human height or crop yield. The rationale and the steps of this approach, which is generally referred to as 'genetical genomics' [3,4] or 'the genetics of gene expression' [5,6], are outlined in Figure 1.

Mapping expression phenotypes to *cis* and *trans* positions

Armed with the expression data from 6215 yeast genes and linkage analysis with a 3114-marker map, 2294 expression phenotypes could be mapped to the yeast genome. A simple non-parametric, single-point mapping method was sufficient to map these phenotypes. In view of the statistical power of such analyses in a population of 86 individuals [7], this large number indicates that a major proportion of the variation in expression levels is of genetic origin and is attributable to genes with relatively large effects on variation. Figure 2 summarizes the main quantitative results obtained in the study.

Of the 2294 expression phenotypes that mapped to the yeast genome, only 25% co-localized with the corresponding structural gene. In these cases, the regulation of expression is probably due to the gene sequence itself (including the flanking regulatory regions). An intriguing observation is that 75% of the gene expression phenotypes do not map to the region containing the structural gene. An essentially similar conclusion is reached when genes are first clustered (hierarchical) on the basis of similarity of their expression profile before linkage analysis. In a genetic study of the mouse brain proteome, it was also observed that variations in protein forms frequently map to positions that are different from the position of the corresponding gene [8].

If the expression phenotype, which is the variation in the expression of a gene (for example, gene X), maps to a different position in the genome (for example, position Y), then the preferred interpretation is that position Y contains a *trans*-acting modulator of gene X. The yeast data indicate that the majority of the genes (X) that show linkage based on the analysis of expression levels are affected by variations in genes located elsewhere in the genome (Y). A reasonable assumption would be that most or all Ys point to transcription factors. However, this assumption might be false [1]. Previously, the transcription factor Hap1 was identified as the cause of expression variation in yeast strains [2]. Through different approaches (Figure 2), Yvert and coworkers now show that

Corresponding author: Ritsert C. Jansen (r.c.jansen@cs.rug.nl).

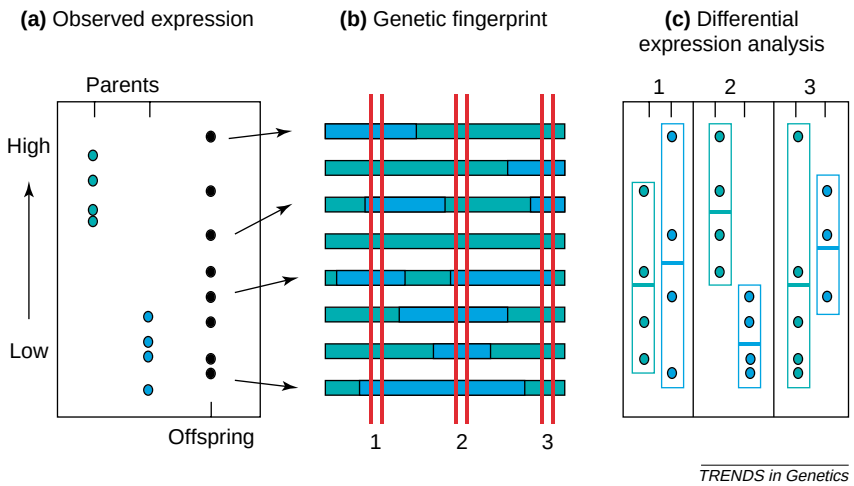

Figure 1. Combining genotyping and expression profiling to study regulatory variation: an outline of the steps in genetical genomics [3,4]. For the sake of clarity, the approach is presented for eight segregants; the actual study used a much larger population of 86 yeast segregants [1]. **(a)** Expression of each gene that was represented on the yeast genome microarray was measured for both parental strains (each in quadruplicate) and once for each segregant in controlled laboratory conditions. The data of one gene are shown as an example. **(b)** The genotype of all offspring is characterized with the help of molecular markers. Green and blue indicate the origin of genome segments from the respective parental strains. Three marker positions are indicated as examples. **(c)** The expression phenotype of segregants with the green allele is compared with the expression phenotype of the segregants with the blue allele, for each of the three markers. The horizontal line is the mean expression level. The difference is (most) significant at position two (based on standard non-parametric tests or the *t*-test). This is taken as evidence for the presence of a regulator locus near position two; the expression phenotype of the gene maps to the region of position two. The expression phenotype shows *cis*-linkage, when the actual structural gene is also located at, or in the vicinity of, position two, otherwise the expression phenotype shows *trans*-linkage. In this example, the detected regulator locus explains the observed expression difference between the parents: segregants that carry the green-marker allele at position two show high expression (green parent), and segregants that carry the blue-marker allele at position two show low expression (blue parent).

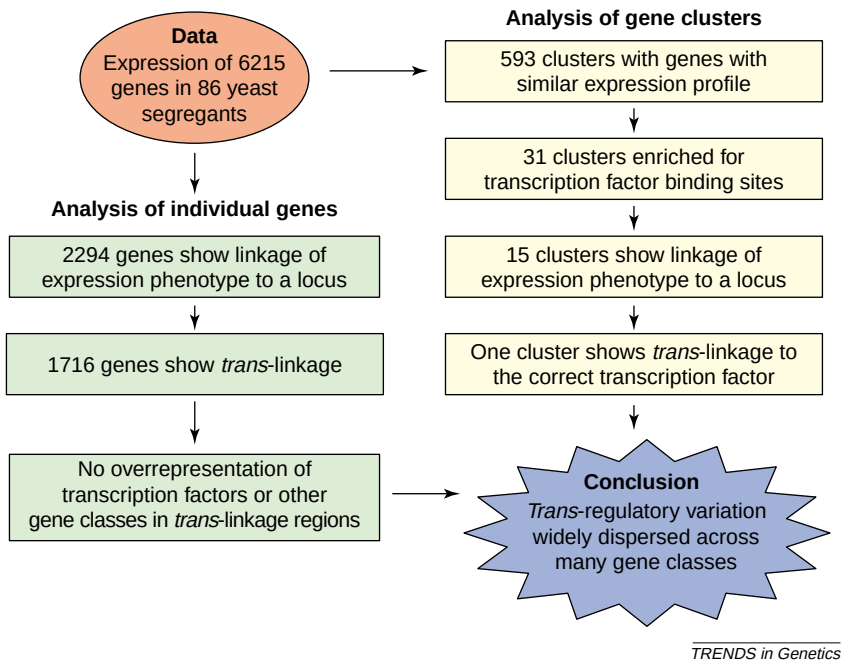

Figure 2. A summary of the main quantitative results obtained in the study by Yvert and colleagues. [1]. The expression data of 6215 genes in 86 segregants were examined at the level of individual genes (green) and at the level of clusters of genes with similar expression profiles (yellow). Linkage analysis was conducted to identify genes or clusters that showed *trans*-linkage (see Figure 1 for basic methodology). From the 6215 genes analyzed, the expression profile of 2294 genes showed linkage to some region of the yeast genome and the expression profile of 1716 genes showed *trans*-linkage. Analyses of the genome regions showing *trans*-linkage for the presence of transcription factors (using a list of 123 known transcription factors) did not reveal any over-representation of transcription factors. Similar strategies were used to investigate the over-representation of other gene ontology classes. Hierarchical cluster analysis identified 593 clusters of more than one gene with similar expression profiles. These clusters were analyzed for the presence of 113 transcription factor-binding sites to investigate whether genes belonging to the same cluster have binding sites in common. In total, 31 clusters showed significant enrichment for binding sites of one or more transcription factors, 15 of which showed *trans*-linkage. However, only one of these 15 clusters linked to the 'correct' transcription factor (i.e. this cluster showed significant enrichment for binding sites of the transcription factor that was detected by linkage analysis). The analyses of all individual genes and clustered groups indicate that, in addition to transcription factors, there are other major players in *trans*-acting regulatory variation.

most regulatory variation that is observed in this yeast population is not caused by the genetic variation in the DNA sequence of transcription factors.

Causes of *trans*-regulatory variation

Perhaps the genes that encode transcription factors are not sufficiently polymorphic in this population study. Genes that do not vary between the parents will not generate variability in the offspring. Consequently, genes that do not tolerate variation will not be identified as major regulators in a genetic approach. More crosses and, if necessary, newly generated variability should be used to investigate this issue [6]. If, however, the genes that encode transcription factors are sufficiently polymorphic, then the results imply that the intuitively important role of transcription factors in gene regulation should be reconsidered.

The experiments of Yvert and coworkers using the genetics of yeast suggest that the importance of transcription factors deserves serious reconsideration. Firstly, the expression phenotype of a cluster of 12 different genes mapped close to *AMN1*, encoding a negative regulator of the mitotic exit network, which has an unknown function in transcriptional enhancement. Genetic analyses indicated that the variation was due to a loss-of-function mutation in *AMN1*, resulting from the change of an acidic residue to a hydrophobic residue [1]. Secondly, the expression level of a given cluster of seven genes mapped to *GPA1*, encoding a G-protein subunit that is coupled to pheromone receptors [1]. Analysis of the parental alleles and comparison with other known yeast sequences suggested that a single serine to isoleucine mutation at amino acid 469 of Gpa1 in the laboratory strain was responsible. Site-directed mutagenesis of *GPA1* in the yeast laboratory strain confirmed the importance of this single amino acid conversion [1].

Gpa1 is a signalling protein in a cascade that results in the activation of a transcription factor. Thus, rather than a transcription factor being the causal agent, a minor change in a protein – a change that supposedly only changes its activity – is responsible for expression variation at different genes in the offspring. This implies that variation upstream in a particular pathway can effect gene expression downstream in that pathway. Not many approaches could have identified such a minor change as the cause of variation. For example, it is unlikely that high-throughput transcriptomics or proteomics on the parental strains alone would have come to such a conclusion.

The three examples discussed above, *HAP1* [2], *AMN1* and *GPA1* [1] suggest that the cause of *trans*-regulatory variation is dispersed over different categories of genes. This is supported by gene ontology analysis: none of the ontology classes are over-represented in the *trans*-linkage regions (Figure 2). Binding-site analysis of transcription factors also supports this conclusion. Of the 15 out of 31 clusters that showed a significant enrichment for the binding sites of one or more transcription factors, only one cluster mapped to a locus in the genome that contained the transcription factor for which the enrichment was observed [1].

Future analyses

The amount of work, data and analyses in the single paper by Yvert and coworkers [1] is impressive, if not possibly somewhat discouraging for laboratories with fewer resources. Moreover, nearly each step in the bioinformatics analysis pipeline is based on statistical assumptions and approaches, part of which can be improved on in the future. Analyses such as multipoint linkage for epistatically interacting regulatory factors might provide additional biologically relevant insights from the available data. Therefore, it is fortunate that all expression data are available in the Gene Expression Omnibus (GEO) repository (accession number GDS464/465; http://www.ncbi.nlm.nih.gov/geo/) for extended analysis. Genotypes of all segregants are, or will be, available on the Kruglyak group website (http://www.fhcrc.org/labs/kruglyak/Data/).

The merger of genetics and genomics had previously identified the existence of hotspots of *trans*-acting loci [2,9] and now sheds light on important regulators in genetic variation. If a large proportion of expression variation is of genetic origin and is a result of genes with relatively large effects, then mapping expression phenotypes might be more promising than mapping traditional phenotypes. Integrating the results from multiple expression analyses seems a more effective way forward in dissecting the complexity of regulatory networks. Exploiting the results discussed in this article and the forthcoming rich sources of integrated data on genes, markers and phenotypes is likely to provide many more surprises for biologists.

References

1 Yvert, G. *et al.* (2003) Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. *Nat. Genet.* 35, 57–64
2 Brem, R.B. *et al.* (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296, 752–755
3 Jansen, R.C. and Nap, J.P. (2001) Genetical genomics: the added value from segregation. *Trends Genet.* 17, 388–391
4 Jansen, R.C. (2003) Studying complex biological systems using multifactorial perturbation. *Nat. Rev. Genet.* 4, 145–151
5 Darvasi, A. (2003) Genomics: gene expression meets genetics. *Nature* 422, 269–270
6 Vogel, G. (2003) Scientists dream of 1001 complex mice. *Science* 301, 456–457
7 Van Ooijen, J.W. (1999) LOD significance thresholds for QTL analysis in experimental populations of diploid species. *Heredity* 83, 613–624
8 Klose, J. *et al.* (2002) Genetic analysis of the mouse brain proteome. *Nat. Genet.* 30, 385–393
9 Schadt, E.E. *et al.* (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422, 297–302

0168-9525/$ - see front matter
doi:10.1016/j.tig.2004.03.002